

# 固有値問題を用いた カーネル多変量解析(1)

05T4007T 江口 晃

- 概要

一般に、「データの情報を縮約する」という問題の多くは固有値問題として定式化できる。

# カーネル主成分分析

- 高次元空間の縮約方法
  - 低次元空間に射影する
  - いくつかの離散点で代表させる
- 主成分分析
  - データの低次元構造を抽出(次元圧縮・次元削減という)することができる。

# 低次元構造の抽出と情報量(1)

- データを集約する
  - もとのデータの情報量をできるだけ保つような低次元構造を抽出する。

次の二つの等価な基準の最適化を行う

## 低次元構造の抽出と情報量(2)

-低次元に射影したときに、ばらつき(分散)ができるだけ大きくなるようにする。

-縮約したデータをもとのデータの近似とみなしたとき、その近似誤差(二乗誤差)ができるだけ小さくなるようにする。

# カーネル主成分分析と固有値問題(1)

- 簡単な射影

$x = (x_1, x_2)$  という2次元の変数で作られる

$$a_1 x_1 + a_2 x_2^2 + a_3 x_2 + a_4 = 0 \quad (3.3)$$

という式は非線形な曲線であるが、

$$\phi_1(x) = x_1, \quad \phi_2(x) = x_2^2, \quad \phi_3 = x_2 \quad (3.4)$$

という変換を考えると、

## カーネル主成分分析と固有値問題(2)

$$a_1\phi_1(x) + a_2\phi_2(x) + a_3\phi_3(x) + a_4 = 0 \quad (3.5)$$

は  $\phi_1 \phi_2 \phi_3$  の作る3次元空間で見れば平面の方程式である。

- カーネル主成分

高次元の特徴ベクトルに変換してから、通常の主成分分析を行って、低次元の線形部分空間を求める。

# カーネル主成分分析の概念図

# 平均0の場合(1)

-まずは、特徴ベクトル  $\phi(x)$  を1次元の直線上に(直交)射影し、

$$f(x) = \omega^T \phi(x) \quad (3.6)$$

という関数を考える。ただし  $\omega$  は単位ベクトルである。

-簡単のため  $\phi(x)$  のサンプル平均を0とする。

このとき、射影した点のサンプル分散を  $\text{Var}_n[f(x)]$  と書くと、

$$\text{Var}_n[f(x)] = \frac{1}{n} \sum_{i=1}^n (\omega^T \phi(x^{(i)}))^2 \quad (3.9)$$

となる。

## 平均0の場合(2)

-(3.9)を最大化するという問題を解けばよい。

制約付きの最大化問題を解くには、ラグランジュの未定乗数法が使える。

最適化したい目的関数の後ろに制約条件を加えた

$$L(\omega) = -\text{Var}_n[f(x)] + \lambda(\|\omega\|^2 - 1) \quad (3.10)$$

の極値問題を解けばよい。

## 平均0の場合(3)

-導入した  $\alpha$  を使うと、 $f(x)$  のサンプル分散は

$$\text{Var}_n[f(x)] = \frac{1}{n} \alpha^T K^2 \alpha \quad (3.14)$$

と書けるので、 $L(\omega)$  を  $\alpha$  で書き直して式変形すると

$$K \alpha = \lambda \alpha \quad (3.17)$$

となり、グラム行列の固有値問題である。

- $L$  を最小にするのは、 $\lambda$  ができるだけ大きな値をとるとき、すなわち  $K$  の最大固有値のときである。

## 平均0の場合(4)

- 2次元以上の空間に射影

上からM個の固有値を取ってきて、対応する固有ベクトルによって張られる空間に射影を取ればよい。

$$f_i(x) = \sum_{j=1}^M \alpha_{ji} k(x^{(j)}, x), \quad j = 1, \dots, M, \quad \alpha_j = (\alpha_{j1}, \dots, \alpha_{jn})^T \quad (3.19)$$

# 一般の場合

-「分散=二乗平均-平均の2乗」という関係を使ってサンプル分散の値を導くと、

$$Var_n[f(x)] = \frac{1}{n} \alpha^T K J_n K \alpha \quad (3.22)$$

となる。ただし $J_n$ は

$$J_n = I_n - \frac{1}{n} \mathbf{1}\mathbf{1}^T \quad (3.23)$$

という行列である。したがって(3.17)は

$$J_n K \alpha = \lambda \alpha \quad (3.24)$$

という固有値問題になる。

# カーネル主成分分析の問題点と データ依存カーネル

- 主成分分析の特徴

- 分散を最大にする低次元空間を求める。
- 二乗誤差を最小にする情報の圧縮法でもある。

-これらの性質はすべて特徴ベクトルでの話で、もとの空間での分散や二乗誤差とは異なる。

-特徴ベクトルの選び方(=カーネル関数の選び方)によって結果が変化してしまう。

# カーネル関数による違い

# 正定値行列とカーネル関数の等価性

- カーネル関数の性質

- サンプルと同じサイズの任意の正定値行列  $K$  が与えられたとき、特徴ベクトル  $\phi(x_{(i)})$  が存在し、それに対応するカーネル関数の定めるグラム行列が  $K$  になる。

データに応じて正定値行列をうまく設定することが可能となる。

# トランダクション1

- 与えられていない入力の扱い
  - サンプルとして与えられていない入力も考えられるものはすべてグラム行列の計算の際に加えておくとよい。

サンプル点以外の入力情報も学習時に取り込んでおく枠組みのことをトランスダクションという。

ただし、入力 $X$ が多い場合、学習にかかる計算量が大きく増大してしまう可能性がある。

# 類似度と正定値行列

- カーネル関数は特徴ベクトルの内積なので、特徴ベクトルどうしの類似度を表していると考えることができる。
- 与えられた類似度が正定値であれば、「正定値行列とカーネル関数の等価性」により、これをカーネル関数として使うことができる。